基于资源分配网络的小数据集并行集成学习方法 *

张安国 [1]，张树勋 [2,3,4]，朱　巍 [1]，李秀敏 [5]，黄金龙 [6?]

(1. 锐捷网络股份有限公司 锐捷研究院，福州 350002; 2. 中国科学院新疆理化技术研究所，乌鲁木齐 830011; 3. 中国科学院大学，北京 100049; 4. 新疆民族语音语言信息处理实验室，乌鲁木齐 830011; 5. 重庆大学 自动化学院，重庆 400044; 6. 长江师范学院，重庆 408100)

摘　要：机器学习领域中，如何在小规模的训练数据集上获得一个具有稳定的高计算精度的算法模型，一直以来都是一个棘手而富有挑战的问题。从算法模型出发，提出了一种基于扩展卡尔曼滤波器的资源分配网络并行集成学习方法。该集成系统由多个带有扩展卡尔曼滤波器的资源分配网络（RANEKF）组成，并且每个 RANEKF 子网的输入由原始数据集中的输入经过随机权值的修正得到。通过和其他神经网络构成的集成学习算法的实验对比，发现提出的方法在小训练集上拥有更高的计算精度和稳定性。

关键词：资源分配网络；并行集成学习；增量学习；扩展卡尔曼滤波器

中图分类号：TP183

Parallel ensemble learning method based on resource allocating networks for small dataset

Zhang Anguo[1], Zhang Shuxun[2, 3, 4], Zhu Wei[1], Li Xiumin[5], Huang Jinlong[6?]

(1. *Research Institute of Ruijie*, *Ruijie Networks Co*, *Ltd*, *Fuzhou* 350002, *China*; 2. *Xinjiang Technical Institute of Physical & Chemical*, *Chinese Academy of Sciences*, *Urimqi* 830011, *China*; 3. *University of the Chinese Academy of Sciences*, *Beijing* 100049, *China*; 4. *Xinjiang Laboratory of Minority Speech &Language Information Processing*, *Urumqi* 830011, *China*; 5. *College of Automation*, *Chongqing University*, *Chongqing* 400044, *China*; 6. *Yangzte Normal University*, *Chongqing* 408100, *China*)

Abstract: To design a training model with stable computational performance and high accuracy which is applied on a small training dataset has been a difficult and challenging problem in the field of machine learning. This paper proposed a Resource Allocating Networks with Extended Kalman Filter (RANEKFs) based parallel ensemble learning algorithm. The learning system is composed of multiple RANEKF units, and the unit inputs are produced by the original dataset with random initialized weights. Based on the experiment results conducted on a small dataset, it is found that the novel model outperforms the ensemble learning systems constructed by the other artificial neural networks in terms of the computational accuracy and stability.

Key Words: resource allocating network; parallel ensemble learning; incremental learning; extended Kalman filter

0　引言

近些年来，人工智能，特别是机器学习得到了学术界和工业界的广泛关注和应用。机器学习中的训练是指算法模型中的自由参数在一定的数据环境下动态的自适应调整过程，从而使模型具有新的输入/输出计算行为，这也是“学习”的内涵所在。因此，在模型学习过程中，训练数据的数量和质量对模型有很大的影响，因而数据的采集和处理也变得尤为重要。在很多实际工程中，数据的采集是一项非常耗时耗力的工作，很难采集到足够多的数据量或者满足所需精度的样本值。例如一些疾病数据以及航空发动机的故障数据，由于本身的发生率特别低，故其数据量的采集变得非常困难。但是，即使是小数据集本身也是有价值的且可学习的，例如本文人类在学习某个新事物的时候，并不需要大量的样本数据。目前，对于数据集的“小规模”并没有明确的定义。通常，本文从两个层面去衡量数据集的规模是否充足，一个是数据量与特征维度之间的关系，另外一个是数据量与噪声水平之间的关系。如果数据量没有远大于甚至小于特征维度，或者数据噪声水平比较高而数据量并不明

基金项目：重庆市自然科学基金资助项目（cstc2016jcyjA0015）；重庆市涪陵区科技计划项目（FLKJ2015ABB1099）

作者简介：张安国（1990-），男，安徽人，研究员，硕士，主要研究方向为人工神经网络、深度学习、机器学习；张树勋（1992-），男，安徽人，硕士研究生，主要研究方向为物联网技术、机器学习；朱巍（1983-），男，四川人，高级研究员，博士，主要研究方向为多媒体技术、人工智能、机器学习；李秀敏（1982-），女，副教授，博士，主要研究方向为神经可塑性、复杂网络、人工神经网络；黄金龙（1989-），男（通信作者），讲师，博士，主要研究方向为离群检测、聚类分析（20170141@yanu.cn）.

显的多，本文可以认为这些情况下数据规模是比较小的。

在小数据集的问题背景下，如何利用有限的数据集来获得一个较高精度的、稳定的机器学习模型，也成为了目前的研究热点。目前解决小数据集上的模型训练问题，主要方法包括三个方向，第一是从训练数据集本身出发，对数据进行过采样或者欠采样[1,2]，第二是从模型算法上出发，例如半监督学习[3,4]，输入特征预处理[5,6]，第三种是从学习策略出发，例如迁移学习[7,8]，集成学习[9]等。

基于弱学习器集成的机器学习模型，在保证整个模型系统具有良好泛化能力的同时，又能避免小训练样本量带来的过拟合现象。因此集成学习的方法成为了一个热门的研究方向，特别是以神经网络作为集成单元中的弱学习器取得了不少的研究成果[10~14]。对于这些集成的小规模神经网络单元，其网络结构和模型复杂度对最终系统计算精度的影响尤为明显，然而目前还没有明确的方法来计算在给定大小的训练集下，如何设计某种神经网络的规模，使其计算性能最优、泛化性最优以及能避免过拟合问题。除了一些经验值以外，很难根据训练样本的数量和复杂度，设计出一个最优的网络模型。因此，能够随着训练集的不断扩张和复杂度的递进而动态改变模型结构和参数的方法——增量学习也被提出来[15~17]。

资源分配网络（resource allocating network，RAN）是一种增量学习的实现模型，其网络隐含层节点的数量能够根据训练样本的复杂度进行动态增加，实现更高的计算能力。因此 RAN 在工业界也得到了广泛应用，例如火电机组控制中的热工对象辨识[18]，自然语言处理中的文本分类[19]，医疗健康领域中的阿尔茨海默病检测[20]，工业生产中的机器人操控[21]等等。另一方面，在学术界，对于 RAN 网络的改进模型也不断被提出来，例如能够防止隐含层节点过快增长的最小 RAN 网络（minimal RAN，MRAN）[21,22]，带有噪声和方差估计的扩展卡尔曼滤波器的 RAN 网络（RAN with extended Kalman filter，RANEKF）[18,19] 等。RANEKF 方法利用扩展卡尔曼滤波器代替最小均方算法进行网络自由参数调整，从而提高了网络的收敛性能，加快其学习速度，有利于在小数据集上快速得到一个稳定可用的网络模型。

在本文中，结合 RANEKF 的优点，提出了一种基于 RANEKF 网络的并行集成学习（parallel ensemble learning，PEL）方法——RANEKF-PEL，同时，为了使每个 RANEKF 子网充分利用训练数据集所有样本数据提供的信息，并且在不同子网间保持一定的差异性，因此对每个 RANEKF 子网的输入赋以不同的随机生成的输入权值向量。本文使用 MNIST 手写字符数据库作为实验测试对象，然而与绝大部分已有研究不同的是本文仅仅使用很小一部分训练数据来训练本文的集成学习系统，并比较了在不同的训练数据集规模，以及不同的集成大小的数据集下 RANEKF-PEL 与其他神经网络集成学习系统的计算精度。实验结果表明，本文所提出的方法具有更高的分类精度和计算稳定性。

1 相关工作

1.1 集成学习

集成学习是指利用多个弱学习器的协作来解决同一个问题。系统的输出由这些弱学习器在当前样本下的输出共同决定。集成学习能够显著提高学习系统在小数据训练集上的计算精度和鲁棒性。目前常用的集成学习方法有 Boosting 和 Bagging [23]。

Boosting 方法的主要思想是在R轮的训练过程中，跟据此轮的训练误差去调整下一轮训练中，每个样本的影响权重。给定一弱学习器h，以及训练样本数据集$< \boldsymbol{X},\boldsymbol{Y} > = \{< x_1,y_1 >,< x_2,y_2 >,\ldots,< x_n,y_n >$，$n$为训练样本量。初始时每个样本赋以相等的权重$1/n$。在$R$轮训练中，每轮训练都对训练失败的样本赋予较大的权重，使其在随后的训练中获得更多的关注。

Bagging 方法的主要思想则是在多轮学习中，每轮从样本数据集$< \boldsymbol{X},\boldsymbol{Y} > = \{< x_1,y_1 >,< x_2,y_2 >,\ldots,< x_n,y_n >$中随机抽取一定量的训练数据来训练集成中所有的弱学习器。集成系统的最终输出由弱学习器投票产生。对于回归问题，则是对所有弱学习器的输出求均值。

1.2 基于 RAN 的集成学习

文献[24]针对单个 RAN 网络稳定性差，易受到样本质量和训练顺序影响的问题，提出了一种由多个最小 RAN 网络单元组成的平行计算结构，其中每个 RAN 网络只使用单一类别（模式）的样本数据进行训练，该方法能够有效避免 RAN 网络的隐含层节点快速增长带来的网络规模过大，易出现过拟合现象的问题。但是由于仅使用样本数据的一小部分来训练单个的 RAN 网络，因此对总的训练数据量需求比较大，无法应用于小样本数据集上，并且对于类异或问题，该方法的有效性会大幅降低，且无法应用于回归拟合问题上。

文献[25]提出了一种基于带有长时记忆（long-term memory，LTM）的 RAN 集成学习方法，并使用 AdaBoost.M1 算法根据当前 RAN 网络的输出误差来计算下一个 RAN 网络的输入样本权值。作者表明了该方法的计算精度优于单一分类器的算法模型。然而，由于该方法训练后一个 RAN 分类器需要使用当前 RAN 的输出误差，在每个时间点只能训练单个的 RAN 网络，无法并行训练所有的 RAN 集合，因此需要消耗大量的训练时间。除此之外，由于 RAN 网络内部并没有随机的初始值，该方法无法在训练初期有效地扩大集合中每个 RAN 之间的差异性，因此其训练效率也不高。

文献[26]提出了一种基于 Bagging 技术的 RAN 网络集成方法，并且为了保证所集成的 RAN 网络具有一定的差异性，每个 RAN 网络只是用$N-1$个类别的样本数据进行训练（N为总共的样本标签的类别数量）。对于集成中的某个 RAN 子网络，若输入的样本与该网络中最近的节点之间的距离大于某一阈值，则表明输入的样本标签不属于该网络训练时的标签空间。对于每个 RAN 子网络，该阈值的选取变的尤为重要，并且该阈值高度依赖于所使用的$N-1$个类别的训练样本，以及剩下的 1 个

类别的训练样本的特征值分布。因此，该阈值的选取将会变得非常艰难和繁琐，甚至无法得到。

综上所述，目前 RAN 网络集成学习方法在小数据集训练样本的问题上表现的并不好，或者需要的训练时间较长，抑或是训练效率不高。

针对前述算法所存在的问题，本文提出了学习算法——RANEKF-PEL，主要具有以下特点：a）算法易构建，训练简单；b）集成单元的全数据集训练；c）动态调整网络结构的复杂度；d）集成单元的并行训练可行性。这些特点很大程度上避免了前述算法所描述的问题，因此本文的算法具有更优秀的性能和适用能力。

2 RANEKF-PEL 方法

本文提出的基于 RANEKF 网络的集成学习方法如图 1 所示，其中输入样本为M维向量$x \in \mathbb{R}^M$，第$i(1 \le i \le p)$个 RAN 分类器的样本输入权值$w_{in}^{(i)} \in \mathbb{R}^M$为随机初始化得到的，并且在后续的训练过程中是固定不变的。整个框架的最终输出$y \in \mathbb{R}^L$由 RAN 网络集合中每个 RAN 分类器的输出$y(i) \in \mathbb{R}^L, (1 \le i \le p)$投票产生。

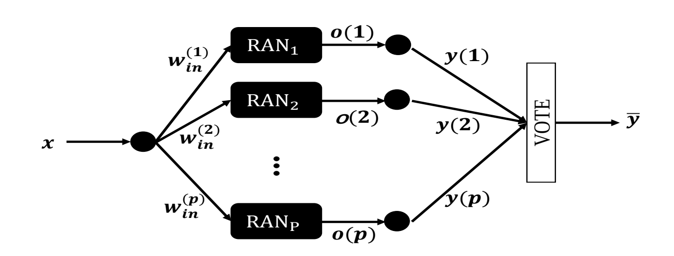


图 1 使用 RAN 作为弱分类器的并行集成学习框架

由于该框架使用 Bagging 集成方法，其中每个 RANEKF 学习器可以采用并行训练的方式，因此仅需很少的时间便能完成整个训练过程。

2.1 网络模型

如图 2 所示的三层结构的 RAN 神经网络，设输入、输出层维度分别为M, L，隐含层的节点数量N可能会在训练过程中，随着输入样本的新颖性发生递增。对于T个训练样本的输入集合$X = \{x_1, x_2, \dots, x_T\}$，其中样本$x_i$包含$M$维特征，即$x_i = \left[x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(M)}\right]' \in \mathbb{R}^{M\times 1}$，隐含层中第$j$个节点的中心为$c_j = \left[c_j^{(1)}, c_j^{(2)}, \dots, c_j^{(M)}\right] \in \mathbb{R}^{M\times 1}$，节点中心宽度为$\sigma_j = [\sigma_j^{(1)}, \sigma_j^{(2)}, \dots, \sigma_j^{(M)}]$。输出层的偏置量$b = \left[b^{(1)}, b^{(2)}, \dots, b^{(L)}\right]'$，对应于输入量$x_i$的网络输出为$y_i = \left[y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(L)}\right]'$。

在计算过程中，隐含层第j个节点的状态为

$$\Phi_j(x_i) = \exp\left(\frac{\|x_i - c_j\|}{\|\sigma_n\|}\right) = \exp\left(\sum_{j=1}^{M} \frac{\left(x_i^{(j)} - c_j^{(j)}\right)^2}{\left(\sigma_j^{(j)}\right)^2}\right) \quad (1)$$

第l个输出层节点的输出为

$$o_i^{(l)} = f(x_i) = b^{(l)} + \sum_{j=1}^{N} w_{lj}\Phi_j(x_i), l = 1,2, \dots L$$

其中：w_{lj}表示隐含层第j个神经元节点到输出层第l个节点之间的连接权值。

对于多分类问题，使用 softmax 算法进行分类，

$$o_i^{(l)} \leftarrow \frac{o_i^{(l)}}{\sum_{j=1}^{L} o_i^{(j)}} \quad (2)$$

最终分类输出为

$$y_i^{(l)} = \begin{cases} 1, \text{if } o_i^{(l)} = \max\limits_{j=1,2,\dots,L}\left(o_i^{(j)}\right) \\ 0, \text{otherwise} \end{cases} \quad (3)$$

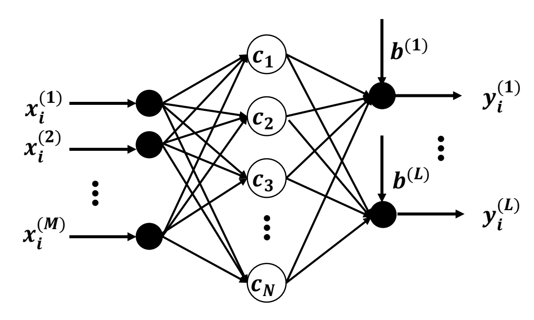


图 2 资源分配网络（RAN）的结构示意图

2.2 网络训练

在初始情况下，RANEKF 网络的隐含层神经元数量为 0。在训练过程中，每个样本数据对$< x_i, d_i >, 1 \le i \le T$，（$d_i$为样本$x_i$的标签值）输入到 RAN 网络时，网络首先计算对应于x_i的实际输出，通过与期望标签d_i的比较判断该样本数据对$< x_i, d_i >$是否满足新颖性条件。如果满足，那么需要在 RAN 的隐含层增加新的神经元节点，反之如果不满足，则需要调节 RAN 的连接权值、节点中心和中心宽度等参数。

具体过程为：

a)初始化 RAN 网络参数。

b)计算样本x_i与当前N个隐含层神经元节点的最近邻欧氏距离：

$$dist(x_i) = \min_{1 \le j \le N}\left\|x_i - c_j\right\| \quad (4)$$

c)计算样本x_i的实际网络输出与期望标签值d_i之间的误差量为

$$err(x_i) = |y_i - d_i| \quad (5)$$

d)如果满足新颖性条件，即$dist(x_i) \ge \delta_{dist}$，并且$err(x_i) \ge \delta_{err}$，那么在 RAN 的隐含层中增加新的节点并初始化，否则对网络的参数进行调整。这里，δ_{dist}和δ_{err}分别为中心节点距离和输出误差的新颖性判定阈值。

e)增加新的隐层节点时，有隐层节点数量：

$$err(x_i) = |y_i - d_i| \quad (6)$$

$$N \leftarrow N + 1$$

新隐层节点的中心：

$$c_N = x_i \quad (7)$$

隐层节点的中心宽度：

$$\sigma \leftarrow [\sigma\,; \lambda \cdot dist(x_i)] \quad (8)$$

网络的输出权值：

$$w_{out} \leftarrow [w_{out}\,; err(x_i)] \tag{9}$$

f)调节网络参数时，根据扩展卡尔曼滤波方法，令

$$\boldsymbol{\Theta} = \left[b, w_{out}^1, c^{1^T}, \sigma^1, \cdots, w_{out}^N, c^{N^T}, \sigma^N\right]^T \tag{10}$$

为所有需要调整的 RAN 网络参数向量，参数更新公式为

$$\boldsymbol{\Theta}(i) = \boldsymbol{\Theta}(i-1) + \boldsymbol{k}(i)err(i) \tag{11}$$

其中增益向量$\boldsymbol{k}(t)$为

$$\boldsymbol{k}(i) = \boldsymbol{P}(i-1)\boldsymbol{B}(i)[R(i) + \boldsymbol{B}^T(i)\boldsymbol{P}(i-1)\boldsymbol{B}(i)]^{-1}$$

这里$\boldsymbol{B}(i) = \nabla\boldsymbol{\Theta} f(x_i)$为函数$f(x_i)$在$\boldsymbol{\Theta}(i-1)$处关于$\boldsymbol{\Theta}$的梯度向量，$R(i)$为测量噪声的方差。协方差矩阵$\boldsymbol{P} \in \mathbb{R}^{z\times z}$ $(z = L + N \times (M + L + 1))$更新方法为

$$\boldsymbol{P}^{z\times z}(i) = [\boldsymbol{I}^{z\times z} - \boldsymbol{k}(i)\boldsymbol{B}^T(i)]\boldsymbol{P}(i-1) + q_0\boldsymbol{I}^{z\times z} \tag{12}$$

这里$\boldsymbol{I}^{z\times z}$为单位矩阵，正标量$q_0$决定梯度向量上所允许的随机步长。需要注意的是当添加新的隐含层节点后，矩阵$\boldsymbol{P}$将变化为

$$\boldsymbol{P}(i) = \begin{pmatrix} \boldsymbol{P}(i-1) & 0 \\ 0 & p_0\boldsymbol{I}^{z_1\times z_1} \end{pmatrix}, \tag{13}$$

其中：p_0时初始参数的不确定性估计，z_1为 RAN 中添加一个新隐含层节点所增加的参数数量，即$z_1 = M + L + 1$。

g)循环步骤 b)~f)，直到$i = T$，即所有的样本都完成训练。

3 实验结果

本文使用 MNIST 手写字符库作为实验数据集。MNIST 包含了由 0~9 十个数字构成的 60 000 组训练数据以及 10000 组测试数据，其中每组数据均为 28*28 像素构成的图像。为了验证提出的方法在小数据集上的有效性，仅会使用训练数据集中的很小一部分作为实验的训练样本集。

同时，为了说明本文提出的方法具有更高的计算精度，使用了多种弱分类器所构成的集成学习方法与本文提出的方法进行对比，这些分类器包括前馈全连接神经网络（feed-forward fully-connected neural networks, FCNs），卷积神经网络（convolutional neural networks, CNNs），深度信念网络（deep belief nets, DBN），以及稀疏自动编码机（sparse auto-encoders, SAEs）。这些分类器已经被成功广泛地用于 MNIST 手写字符识别问题中，在完整的 MNIST 数据训练集上，这些网络都获得了 97%以上的分类准确率，说明这些网络都拥有良好的学习能力，具有对比价值。

比较了几种集成网络在不同的训练数据集规模，以及不同的弱学习器集成规模下的分类精度。

3.1 训练数据集规模

本文测试了在集成规模为 20 的情况下，几种集成学习在训练集规模从 50~2000 下的分类精度，其中每个手写数字类别（0~9）的训练样本数量为

$$N_i = \frac{S}{10}, \qquad i = 0,1,2,\ldots,9$$

其中：S为训练集规模，即保证在小规模训练集中，样本类别的数量分布均匀。

实验结果如图 3 所示。可见在极端小的训练样本规模时，每个类别的样本量仅有 5 个，此时的 5 种模型分类结果对比中，仅有本文提出的 RANEKF-PEL 的精度大于 50%。并且随着训练数据集的增加，几种集成学习方法的分类精度的均值都逐渐增大，并且分类精度的波动都逐渐减小。其中本文提出的 RANEKF 集成学习方法在同等训练数据集规模下拥有最高的分类精度，以及更小的精度波动。

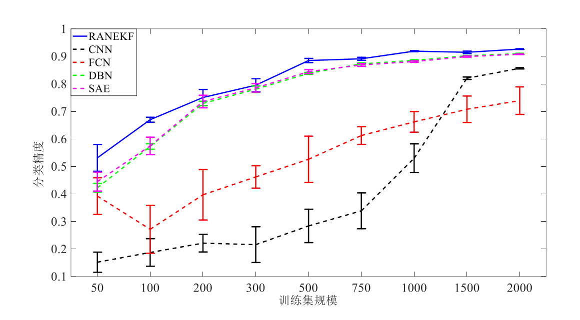


图 3 不同训练集规模下，几种集成学习方法的分类精度和精度波动对比

3.2 集成规模

集成规模的大小对系统的输出精度和稳定性有着非常重要的影响。通常情况下，越小的集成规模的学习，其精度和稳定性也越小。

实验中，对比了不同集成大小（1，3，5，7，10，12，15，18，20）情况下，在训练数据集大小为 100 时的分类输出精度和波动。图 4 中，RANEKF-PEL 方法的分类精度始终能够保持在 50%以上，也表明了本文提出的方法在不同集成规模下均拥有最高的精度和非常小的精度波动，即拥有很高的计算稳定性。

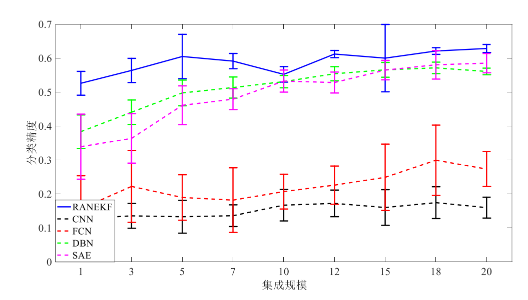


图 4 不同集成规模下，几种集成学习方法的分类精度和精度波动对比

3.3 隐层节点的变化

如图 5 所示，在训练的起初阶段，RAN 的隐含层节点数量非常少，此时网络的输出精度较差，隐含层节点因此增加的速度比较快。而随着 RAN 网络规模的不断增长，网络对复杂模式的计算能力不断增强，不需要再过多地添加新的节点。当隐含层节点数量达到 50 至 70 之间时，网络已经拥有较好的计算能力，因此隐含层的节点数量渐渐趋于某个固定值，网络的

结构状态也渐渐趋于稳态。

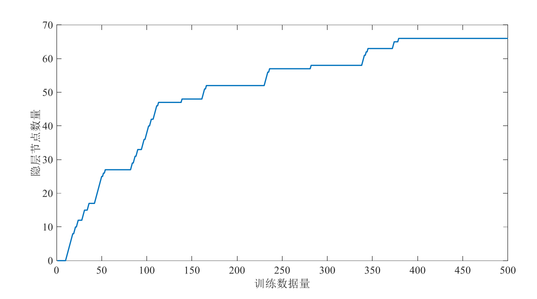


图 5 增量训练过程中，RAN 的隐含层节点变化情况

4 结束语

为了在小规模训练数据集上获得较高的计算精度和计算稳定性，本文提出了一种基于带有扩展卡尔曼滤波器的资源分配网络的并行集成学习方法（RANEKF-PEL），该方法中每个资源分配网络（RAN）单元的输入信号都是原始样本数据各个维度特征的加权信号。通过与其他多个类型的人工神经网络单元所构成的集成系统的实验对比，本文提出的 RANEKF-PEL 方法能够在小数据集上训练后，能够获得最高的计算精度和鲁棒性能，是一种有效可行的机器学习模型。

参考文献：

[1] Chen H Y, Li D C, Lin L S. Extending sample information for small data set prediction [C]// Proc of International Congress on Advanced Applied Informatics. 2016: 710-714.

[2] Li D C, Liu C W. Extending attribute information for small data set classification [J]. IEEE Trans on Knowledge and Data Engineering, 2012, 24 (3): 452-464.

[3] Aydemir M S, Bilgin G. Semisupervised hyperspectral image classification using small sample sizes [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14 (5): 621-625.

[4] Meng Jianjun, Sheng Xinjun, Zhang Dingguo, et al. Improved semisupervised adaptation for a small training dataset in the brain-computer interface [J]. IEEE Journal of Biomedical and Health Informatics, 2014, 18 (4): 1461-1472.

[5] Da Silva I B V, Adeodato P J L. PCA and Gaussian noise in MLP neural network training improve generalization in problems with small and unbalanced data sets [C]// Proc of IEEE International Joint Conference on Neural Networks. 2011.

[6] Sasikala S, Appavu S, Balamurugan A, et al. An efficient feature selection paradigm using PCA-CFS-Shapley values ensemble applied to small medical data sets [C]// Proc of International Conference on Computing, Communications and Networking Technologies. 2013.

[7] Wang Tian, Chen Yang, Zhang Mengyi, et al Internal transfer learning for improving performance in human action recognition for small datasets [J]. IEEE Access, 2017, 5: 17627-17633.

[8] Li Zengxi, Song Yan, Mcloughlin I, et al. Compact convolution neural network transfer learning for small-scale image classification [C]// Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2016.

[9] Rani T S, Soujanya P V. An ensemble method using small training sets for imbalanced data sets: Application to drugs used for kinases [C]// Proc of International Conference on Contemporary Computing. 2013.

[10] Cheng Zezhou, Yang Qingxiong, Sheng. Bing Colorization using neural network ensemble [J]. IEEE Trans on Image Processing, 2017, 26 (11): 5491-5505.

[11] Zhou Zhihua, Jiang Yuan. NeC4. 5: neural ensemble based C4. 5 [J]. IEEE Trans on Knowledge on Data Engineering, 2004, 16 (6): 770-773.

[12] Yu Shicai, Xia Giurong. Neural network ensemble method based on improved sort learning algorithm [C]// Proc of International Forum on Information Technology and Applications, IEEE, 2010.

[13] Liu Yong. Create weak learners with small neural networks by balanced ensemble learning [C]// Proc of IEEE International Conference on Signal Processing, Communications and Computing. 2011.

[14] Lima T P F, Ludermir T B. Differential evolution and meta-learning for dynamic ensemble of neural network classifiers [C]// Proc of IEEE International Joint Conference on Neural Networks. 2015.

[15] Schaal S, Atkeson C G. Constructive incremental learning from only local information [J]. Neural Computation, 1998, 10 (8): 2047-2084.

[16] Gao Yaozong, Zhan Yiqiang, Shen Dinggang. Incremental learning with selective memory (ILSM): towards fast prostate localization for image guided radiotherapy [J]. IEEE Trans on Medical Image, 2014, 33 (2): 518-534.

[17] Ristin M, Guillaumin M, Gall J, et al. Incremental learning of random forests for large-scale image classification [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2016, 38 (3): 490-503.

[18] 杨世忠，吕剑虹. 基于最小资源分配网络的热工对象辨识 [J]. 热能动力工程, 2007, 22 (1): 91-95.

[19] 何晓亮，宋威，梁久帧. 基于资源分配网络和语义特征选取的文本分类 [J]. 计算机功课与科学, 2014, 36 (2): 340-346.

[20] Mahanand B S, Suresh S, Sundararajan N, et al. Alzheimer's disease detection using a self-adaptive resource allocation network classifier [C]// Proc of IEEE International Conference on Neural Networks. 2011.

[21] Corradini M L, Fossi V, Giantomassi A, et al. Minimal resource allocating networks for discrete time sliding mode control of robotic manipulators [J]. IEEE Trans on Industrial Informatics, 2012, 8 (4): 773-745.

[22] CIabattoni L, Grisostomi M, Loppoliti G, et al. On line solar irradiation forecasting by minimal resource allocating networks [C]// Mediterranean Conference on Control & Automation. 2012.

[23] 周志华，陈世福. 神经网络集成 [J]. 计算机学报, 2002, 25 (1): 1-8.

[24] Lim W S, Rao M V C. Effective sonar target classification via parallel structure of minimal resource allocation network [J]. International Journal

of Computational Intelligence, 2008, 18: 1109-1116.

[25] Kidera T, Ozawa S, Abe S. An incremental learning algorithm of ensemble classifier systems [C]// Proc of IEEE International Joint Conference on Neural Networks. 2006.

[26] Ji Genlin, Ling Xiaohan, Cheng Xueyun. Novel distributed intrusion detection method based on neural network ensemble [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2007, 39 (2): 231: 235.